DOCUMENT RESUME

ED 122 258 CS 002 620

AUTHOR Askov, Eunice N.
TITLE Criterion-Referenced Strategies for Diagnosis.
PUB DATE 76
NOTE 10p.; Paper presented at the Annual Meeting of the International Reading Association (21st, Anaheim, California, May 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Tests; *Criterion Referenced Tests; Diagnostic Tests; Elementary Education; Individualized Instruction; *Individualized Reading; *Reading Achievement; *Reading Diagnosis; Reading Tests; Test Bias

ABSTRACT

Standardized achievement tests may not be an accurate diagnostic tool for measuring individual pupil achievement in reading. Criterion-referenced tests, designed to measure specific goal mastery, can provide valuable information that standardized forms cannot. In addition, criterion-referenced tests give teachers the diagnostic tools needed to individualize instruction and increase instructor effectiveness in educational decision-making. Caution must be exercised to ensure that testing reflects the instructional emphases of the school's reading program, and that objectives are written at an appropriate level of specificity. Issues of teacher support, test reliability, and integration of separate skills with sustained reading must also be considered. (RS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED122258

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Eunice N. Askov
The Pennsylvania State University
178 Chambers Building
University Park, PA 16802

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Eunice N. Askov

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

Criterion-Referenced Strategies for Diagnosis

(Paper given as part of the symposium
Diagnostic-Prescriptive Reading
Instruction Strategies)
May 12, 1976, 2:00-4:45 p.m.
Disneyland Hotel-N. Ballroom #2

CS002620

Criterion-Referenced Strategies for Diagnosis

The traditional method for assessing pupil achievement and growth has been the use of results from standardized achievement tests. But this approach has not proved satisfactory for several reasons. Subtests that might be used diagnostically generally lack sufficient reliability to permit subtest results to be used in this manner. Consequently, statistically meaningful scores result from larger tests that are global measures of reading ability rather than discrete measures of specific skill attainment. Teachers often do not find these results useful since the diagnostic value is limited. Furthermore, schools located in low socio-economic areas tend to serve children who score lower on standardized achievement measures than those attending suburban schools (4). Are these tests, then, really testing achievement or something else? Even within a school building, if children are homogeneously grouped, teachers of the low ability groups are at a disadvantage if their pupils' scores are compared to those attained by pupils in higher ability groups. Even if gain scores are used, children of lower ability are still penalized by their slower growth. Children who speak a dialect or whose first language is not English are at a further disadvantage on these formal testing measures (3).

Schools, teachers, curricula, etc. seem to have little impact when standardized achievement tests are used as measures of pupil achievement. Data from the Coleman Report (1), which were reanalyzed by Mayeski et al. (7), reveal that only five percent of the achievement variance can be uniquely attributed to school variables. The Mayeski report states:

> When all groups were combined, and Racial-Ethnic Group Membership was considered as an aspect of FB (Family Background), as much as half of the total differences among students in their ACHV (Achievement Index, a composite of achievement test scores) could be explained by FB. (p. 117)

Therefore, standardized achievement test scores do not seem particularly useful to the teacher since their results do not reflect specific learnings that might have been gained during a school year. It was for this reason that the concept of criterion-referenced testing was developed.

First an objective is written or selected; the objective should state specifically what observable behaviors are expected upon mastery of the objective and the conditions under which the behaviors are to be demonstrated. A criterion performance level, such as 80 percent correct, must also be stated. An example follows: "Given a list of monosyllabic words each containing a single short vowel sound, the child will pronounce the words with 80% accuracy."

Next, items which assess these behaviors are written. The items should directly assess what is stated in the objective. An item, such as meat in the above example, would be unacceptable since it does not fit the conditions stated in the objective, i.e., words containing a single short vowel sound. Nor would asking the child the meaning of a word fit the objective since the objective stated only that the child should be able to pronounce the word.

The advantages of this type of testing for instructional planning are readily apparent. The teacher knows exactly what the child has and has not mastered. For skills not mastered the teacher knows whether the child was close to mastery (scoring, say, 70%) or in need of much work before mastery can be expected (scoring, say, 20%). Instruction is

efficient since it can be focused specifically on skills that children need rather than on skills -- which certain children may have already mastered -- that are taught in connection with a given story in a basal reader. The teacher can compare the diagnostic profiles of a group of children; children with a given skill weakness can be grouped together. If all teachers of children in a given age range (e.g., six to eight years old) assess mastery of children on the same array of skills, then teachers can plan instruction together. One teacher can agree to instruct all children needing a given skill while others can teach other skills. Teachers can capitalize on their strengths and prepare for instruction in a few skills rather than having to offer direct instruction in all skills. The children have the benefit of improved instruction and a variety of teachers.

Some problems may also arise, however, in implementing the diagnostic teaching of skills. If teachers exchange pupils for skill instruction, they must be in constant communication about the progress of each of the pupils. Teachers must schedule time for planning meetings to check pupil progress and coordinate skill instruction. But perhaps the most important consideration is that skill instruction does not become an isolated activity so that children may readily transfer what is learned in skill instruction to sustained reading. In the presentation of new skills, lower level skills should be reviewed. Furthermore, skill instruction should stress application of the skill in context rather than in isolated words. In any kind of directed reading lesson skills should be reviewed and reinforced. Difficulties should be noted and corrected. Use of an informal reading inventory can be helpful in spotting skills that are not being applied in sustained reading as well as checking a child's reading level.

Recently criterion-referenced testing/teaching systems have been attacked for their fragmentation of language processing. The argument is that reading is not learned merely by mastering a series of supposed skills. Reading is instead a holistic process in which "there is no possible sequencing of skills" (5). Samuels (8), however, attempts to reconcile the holistic and skills approaches by arguing that skill acquisition is an essential part of learning to read; early instruction in skills that are commonly recognized can prevent later difficulties in reading. If, then, one takes the view that skill acquisition is one part of learning to read -- as is the assumption underlying most reading programs today -- then criterion-referenced testing/teaching systems can provide useful diagnostic information about children's skill mastery.

What, then, are the important considerations in creating or selecting a criterion-referenced testing system? The objectives are, first of all, of paramount importance. They must be compatible with the instructional emphases of the school's reading program. It is pointless to assess areas that are not being taught. If the objectives are written by teachers, then they are more likely to reflect current reading instruction. If, however, a commercial system is to be chosen, the objectives must be carefully scrutinized. Do they state the important aspects of instruction? Is the sequence of objectives similar to the sequence found in the school's instructional program and materials?

Are the objectives written at a mid-level specificity? Objectives that are too general (e.g., the child will read a passage written at the third grade level with at least 98% accuracy in word recognition) are not useful diagnostically because they do not identify the specific difficulties. On the other hand, objectives should not be so specific

that the reading process does become fragmented and that the results are no longer meaningful. When, for example, a system has over four hundred objectives for the primary grades, management of instruction becomes a colossal task and reading tends to suffer the fate of Humpty Dumpty. If objectives are written at a mid-level specificity (e.g., considering short vowel sounds as opposed to the short a, etc.), then assessment results can be useful diagnostically.

A further, very important consideration is whether technical information was gathered during test development. Some criterion-referenced testing/teaching systems (e.g., Wisconsin Design for Reading Skill Development, 9) were extensively field-tested and revised before commercial publication. Others (e.g., Fountain Valley Teacher Support System in Reading, 2) were published without gathering technical data. If a school district plans to develop its own criterion-referenced tests, then it should also plan to gather technical data in order to create effective tests, If tests are not well developed, then the rest of the additional, supportive components (e.g., a management system, prescriptive teaching resources) are of questionable value.

Particulary important among technical data gathered is test reliability. If objectives are written at a very specific level, then practicality would mandate that only a few items could be used to assess each objective. Test reliability simply is not adequate with very brief assessments of an item or two. Objectives written at mid-level specificity may be assessed by more items (say, 15 or 20), yielding more reliable tests.

A final consideration in creating or selecting a criterion-referenced testing/teaching system is the type of teacher support desired. Some commercial systems provide a record-keeping system, such as punch cards

which are notched whenever a child demonstrates mastery of a skill. Another provides a computer printout indicating individual skill mastery or deficiency along with prescriptions for instruction. Other systems, rather than provide specific prescriptions, provide resource files which suggest a variety of activities and materials from which a teacher may choose. In other words, some commercial systems are less prescriptive than others, rather than placing confidence in the teacher's ability to decide what is the most appropriate and effective instruction for individual pupils.

Conclusion

Criterion-referenced tests, whether constructed locally or commercially, can provide valuable diagnostic information. A teacher can plan specific skill instruction, focused on just the skills needed by individual pupils. If a child has already mastered a skill, skill mastery is noted, and he/she can move on to new skills rather than spending time on skills already mastered. Criterion-referenced tests give teachers the diagnostic tools needed in individualizing instruction. The teacher becomes much more effective in instructional decisions. He/she becomes a decision-maker rather than a facilitator of directions found in a teacher's manual.

Criterion-referenced testing, however, should be supplemented with other types of assessment. Standardized achievement tests, while not useful diagnostically, do indicate whether a group of children are making progress in relation to a norm group or a national standard. Informal reading inventories are also useful, indicating how well a child is able to integrate the various skills in a realistic reading task. Attitudes, likewise, should not be forgotten. Knowing how to read is essential, but the desire to read is also important.

A final caution, involving the accountability process, should be mentioned. With criterion-referenced testing it is very easy to determine which children, given instruction in specific skills, have mastered those skills. It, therefore, would be easy to conclude that the teacher whose children demonstrate mastery of more skills is the more effective teacher. While this assumption may be to some extent true, one must keep in mind that criterion-referenced tests even longer ones -- lack the test reliability of standardized achievement tests. Therefore, while criterion-referenced tests may certainly be used to guide instruction, only results from a fairly large group of children should be used in teacher or program evaluation (6).

Keeping limitations of criterion-referenced testing in mind, teachers may effectively use these tests to diagnose skill attainment and to provide instruction focused on skill needs. While criterion-referenced testing has been going on in the classroom for many years, only within about the past ten years was the concept articulated and systems devised to aid the teacher in reading instruction. Now an array of commercial systems exists with varying degrees of teacher support and at differing prices. Teachers and administrators should first analyze their own instructional program to determine what is needed and then seek or develop a compatible system which will provide the most useful information for local teachers.

References

1. Coleman, J. S., et al. Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

2. Fountain Valley Teacher Support System in Reading. Huntington Beach, California: Richard L. Zweig Associates, 1971.

3. Green, R. L. "Comments on Ebel's Defense of Tests and Testing," Phi Delta Kappan, 57 (October 1975 a), 88-89.

4. Green, R. L. "Tips on Educational Testing; What Teachers & Parents Should Know," Phi Delta Kappan, 57 (October 1975 b), 89-93.

5. Johnson, D. D., and Pearson, P. D. "Skills Management Systems: A Critique," The Reading Teacher, 28 (8), May 1975, pp. 757-764.

6. Massad, C. E. "Development and Use of Objective-Referenced Tests for Instructional Purposes." Presentation at the First Eastern Regional Conference of the International Reading Association, 1976.

7. Mayeske, G. W., et al. A Study of the Achievement of Our Nation's Students. Washington, D.C.: U.S. Government Printing Office, 1973.

8. Samuels, S. J. "Hierarchical Subskills in the Reading Acquisition Process." Paper presented at the annual meeting of the International Reading Association, 1975.

9. Wisconsin Design for Reading Skill Development. Minneapolis: National Computer Systems, 1972.